Original Article

Comparison of Cancer Registry Case Ascertainment with SEER Estimates and Self-reporting in a Subset of the NIH-AARP Diet and Health Study

Dominique S. Michaud[a,b], ScD; Douglas Midthune[c], MS; Sigurd Hermansen[d]; Michael Leitzmann[a], DrPH; Linda C. Harlan[e], PhD; Victor Kipnis[c], PhD; Arthur Schatzkin[a], PhD

Abstract: The NIH-AARP Diet and Health Study cohort consists of 567,169 members of the AARP, aged 50-69 years and living in 8 different states, who responded to a mailed questionnaire in 1995-1996. At baseline, most cancer registries included in this large cohort study were not part of the Surveillance, Epidemiology, and End Results (SEER) Program. We undertook a pilot study to determine the completeness of case ascertainment using record linkage to 8 U.S. cancer registries. We conducted a pilot study of 12,000 cohort participants. A number of identifiers were used to match our population to cancer registries. In addition, we mailed a questionnaire to the participants in the pilot study to obtain cancer information directly, and attempted to confirm self-reported cancers by retrieving medical records. Overall and site-specific cancer incidence rates were similar to those calculated by applying SEER rates to the pilot study population. We estimated a minimum sensitivity of 89.2% for case ascertainment when the registry search lags the end of the follow-up period by 4 years. The accuracy of case ascertainment through cancer registries in this cohort is comparable to that of other cohort studies that rely on self-reporting (with medical record confirmation) for case ascertainment, and should provide accurate cancer numbers for the NIH-AARP Diet and Health Study cohort analyses.

Key words: *cancer registry, case ascertainment, record linkage*

Introduction

To date, most validation studies on cancer ascertainment have relied on cancer registries to estimate sensitivity and specificity of self-reported cancers. These studies compared self-reported cancers to cancer registry information, treating cancer registries as the "gold-standard" when determining the quality of the self-reports.[1-5] The North American Association of Central Cancer Registries (NAACCR) has a program to evaluate the accuracy, completeness, and timeliness of cancer registry data, and certifies registries that meet the criteria for the highest standard of data quality.[6] The National Cancer Institute (NCI) Surveillance, Epidemiology, and End Results (SEER) Program's standard for case ascertainment is 98% for their registries.[7]

Cancer registries participating in the SEER Program are expected to provide cancer incidence data within 19 months of the close of a diagnosis year (e.g., 1998 cancer data must be reported to NCI by August 2000). A recent study showed that, depending on cancer site, 88-97% of cancers diagnosed in the SEER registry states are reported to the NCI within a 2-year period.[8] The same study concluded that it would take 4-17 years, depending on the cancer site, for 99% or more of the cancer cases to be reported to the NCI. However, no study has investigated the accuracy of non-SEER registries certified for the highest standard of data quality.

In this paper, we report results from a pilot study of the NIH-AARP Diet and Health Study cohort comparing observed number of cancers to expected numbers, based on SEER registry data. In addition, we examined self-reported cancers from pilot participants to assess the accuracy and completeness of the data obtained from cancer registries (most of which were not SEER registries) used in the NIH-AARP cohort.[9]

Methods

Study Population

The NIH-AARP Diet and Health Study cohort was initiated between 1995 and 1996 when 567,169 members of the AARP, aged 50-69 years, returned questionnaires sent to them in the mail. This cohort was designed to examine the relation of diet with cancer outcomes and to address some methodological issues often encountered in studies on diet and cancer. A 16-page questionnaire was mailed in 1995-1996 to 3.5 million AARP members in 6 states (California, Florida, Pennsylvania, New Jersey, North Carolina, and Louisiana) and 2 metropolitan

"Comparison of Cancer Registry Case Ascertainment with SEER Estimates and Self-reporting in a Subset of the NIH-AARP Diet and Health Study"

[a]Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD. [b]Department of Epidemiology, Harvard School of Public Health, Boston, MA. [c]Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD. [d]Westat, Inc., Rockville, MD. [e]Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD.

Address correspondence to Dominique S. Michaud, ScD; Harvard School of Public Health, 667 Huntington Ave, Kresge 920, Boston, MA 02115. Telephone: 617-432-4508; Fax: 617-566-7805; e-mail: dmichaud@hsph.harvard.edu.

areas (Atlanta, Georgia and Detroit, Michigan). The recruitment areas were selected for having high-quality registries and large AARP memberships. Additional details on the design of this cohort can be found in a previous publication.[9]

Cohort Maintenance

From the onset of the study, the NIH-AARP cohort has been matched periodically to the National Change of Address database maintained by the U.S. Postal Service (USPS) to update address changes. Information on address change is also obtained through USPS processing of undeliverable mail and directly from participants who report address changes in follow-up questionnaires. During the first 3 years of follow-up, 98% of the cohort members either did not move or relocated within 1 of the 8 states in the study. In addition, the cohort data are periodically matched to the U.S. Social Security Administration's Death Master File to find out which participants are deceased.

Cancer Registries

The 8 statewide cancer registries (for 2 metropolitan areas and 6 states) did not belong to the SEER Program at baseline. The NAACCR standards state that cancer registries should contain at least 95% of the expected cases of reportable cancer within 18 months of the close of a diagnosis year.[6] The NAACCR Certification Committee has established criteria for recognizing cancer registries that achieve excellence; these criteria include timeliness and 2 measures of reporting completeness.[6] All cancer registries included in this study were certified by NAACCR for meeting the highest standard of data quality (beginning in 1997).

Pilot Study to Assess Case Ascertainment

Twelve-thousand participants in the NIH-AARP cohort were randomly selected from the 6 states and 2 metropolitan areas of 2 different states (for a total of 8 states); numbers from each area were set to reflect cohort size in those areas. We estimated 12,000 participants were necessary to determine whether cancer rates from this cohort were within the expected rates (using SEER cancer registry incidence rates). The follow-up period for cancer incidence was 1995 through the end of 1998. For example, the sample size of 12,000 provided 80% power to detect a colorectal cancer incidence rate for men and women combined that was 25% lower than the SEER rate.

Matching Participants to Registry Data

Most registries use a number of identifiers to optimize the matching of individuals to their databases. For the pilot study, we had almost complete data on first and last names, date of birth, gender and address, as well as 85% of Social Security numbers, for matching purposes. All of the initial linkages were undertaken between July 2001 and December 2002. In addition, we conducted a linkage of the full cohort, including the 12,000 pilot study subjects, between January and August 2003.

With the exception of 2 states, linkages first took place at the registries and again, independently, at Westat (an employee-owned research corporation). For Pennsylvania, the linkage was carried out only at Westat, and for Florida the linkage was only carried out at the registry. Since the second linkage benefited from a more technically-advanced, automated process and a more thorough review process, we report results from the second linkage for the 12,000 pilot study subjects.

Seven of the 8 registries participating in the NIH-AARP cohort routinely use probabilistic record linkage methods to link cohort records to registry databases (AUTOMATCH® or one of its successors). Probabilistic linkage uses a set of identifiers to predict the probability that records from different datasets are true matches. Controlling match error rates, probabilistic methods minimize the numbers of possible matches requiring review.[10] The Westat program (currently version 1.16) builds multiple compound indexes for alternative patterns of identifiers in subject records, and it uses the indexes to screen each registry record for a 'hit' on any one of the indexes;[11] record pairs screened out as possible matches then go through a series of comparisons that reject those with insufficient degree of similarity to confirm a true match. The Florida registry used the Westat program to select possible matches and reviewed the possible match to determine true matches to cancer reports.

To improve our chances of finding true matches in the cancer registries, multiple records were created for subjects with different sets of identifying information (e.g., 2 mailing addresses). For those subjects who had moved from 1 to another of the 8 states in our study, we included their records to be matched in both states.

Cancer Self-Reports in Responses to Follow-up Questionnaire

Beginning in April 2002, and in parallel with registry matching, follow-up questionnaires were sent to pilot participants who were still alive (n=11,404). Participants were asked whether they had been diagnosed with cancer since 1995 (other than non-melanoma skin cancer). Participants who reported a cancer during that time period were asked to specify the type of cancer, as well as date of diagnosis. Up to 3 cancers could be reported.

Participants who did not respond to the initial mailing were sent a second follow-up questionnaire, and then a postcard inviting them to call a toll-free number and provide information using an automated system. The remaining nonrespondents were contacted by trained interviewers whenever possible (i.e., if the telephone numbers were available). Survey data collection ended in July 2002.

Of the 12,000 pilot participants, 596 were found to be deceased prior to mailing the questionnaires, and 41 were found to be deceased during follow-up. A total of 6,695 questionnaires were returned by mail, 75 were completed through the automated phone system, and another 1,856 questionnaires were completed by phone interviews. Four subjects refused to participate; the remaining 3,033 did not respond to any mailings and were not contacted (no telephone number available or no answer after multiple attempts to contact the individual by phone). Overall, the response rate was 73.3% (questionnaire returns among live participants = 8,326/11,363). Basic characteristics were similar across responders and non-responders (41% female in both groups; 93% White among responders, 90% White among nonresponders; mean age 61.6 and 60.6 for responders and nonresponders, respectively).

Data Analysis

We compared the number of registry cancer cases (matched to pilot study participants) to the number of expected cancer cases based on the SEER Program data, and performed χ^2 tests of the differences. We calculated the expected numbers of cases with the method described by Monson.[12] This method uses birth, entry, and exit dates to determine how much time at-risk a subject spent in each 5-year age group and then uses age-specific incidence rates to calculate overall expected cases. The age-specific incidence rates were calculated using the SEER*Stat program,[13] and were based on the 9 SEER Program registries' data in years 1995-1998 (for whites, by gender).[14] Exclusions from these calculations included: *in situ* cancer cases (except for bladder cancer), cases with missing diagnosis date, cases with diagnosis date after the end of December of 1998, and cases with diagnosis date prior to the baseline questionnaire date.

In addition to comparing the observed (O) vs. expected (E) rates obtained from the cancer registry matching, we conducted a comparison between reports from the registries and self-reported cancers from the follow-up questionnaire. For this comparison, we used the same selection criteria used for the O/E calculations, but limited the analysis to survey respondents.

We estimated the number of cancers missed by the cancer registries (false negatives) by multiplying the percent of cancers confirmed with medical records (only available for a subset of total) by the number of self-reported cancers for which no medical records were available. To this number we added the number of confirmed cancers. This number was used to estimate missed cancers for all sites combined, i.e., confirmed self-reported cancer (confirmed self-reported cancer + cancers reported in registries.

Our analyses are based on a 4-year lag period (the time between the end of follow-up for cancer incidence and the linkage date, i.e., December 1998 and December 2002).

Results

A total of 455 cancers were registry matched to the 12,000 participants between 1995 and 1998. Observed and expected numbers of site-specific cancers, by gender, are reported

Table 1. Number of Registry-detected Cancers in the NIH-AARP Cohort Pilot Study (n=12,000) Compared to Expected Numbers Based on SEER Cancer Rates (baseline through December 31, 1998)

Site	Gender	Observed	Expected	P-value
All	Male	318	336	0.32
	Female	137	157	0.12
Lung	Male	41	55	0.06
	Female	23	25	0.70
Colorectal	Male	38	35	0.67
	Female	14	16	0.61
Prostate	Male	127	120	0.53
Breast	Female	52	52	0.99
Endometrial	Female	10	12	0.63
Ovary	Female	2	6	0.09
Stomach	Male	9	6	0.14
	Female	0	1	0.24
Pancreas	Male	10	7	0.25
	Female	2	3	0.46
Kidney	Male	6	9	0.28
	Female	2	3	0.52
Bladder	Male	19	22	0.57
	Female	6	4	0.26
Esophagus	Male	6	5	0.63
	Female	1	1	0.78

Table 2. Number of Self-reported Cancer by Questionnaire versus Registry-reported Cancer Between Baseline (1995–1996) and December 31, 1998 (any cancer match)

Self-reported	Registry-reported		
	Yes	No	**Total**
Yes	206 (20.5%)	82 (1.0%)	288
No	71 (0.8%)	7967 (95.7%)*	8038
Total	277	8049	8326

*Includes self-reported nonmelanoma skin cancers (n=10); 1 precancer lesion; 1 reported metastatic cancer; and in situ cancers (n=18).

Sensitivity of registry = 206/288 = 72%
Specificity of registry = 7967/8038 = 99%
(assuming self-report is the gold-standard)

in Table 1. Small numbers for site-specific cancers limited our interpretation of those results, but over all cancer sites, the observed rates were not substantially different from the expected rates.

Table 2 summarizes the cancer information obtained from the follow-up questionnaire compared to the results of the cancer registry searches. Cancers reported on the questionnaire were used only if date of diagnosis was provided, and the numbers presented are for any cancer match. Percentages of self-reported cancers not found in the registries were similar across the various cancer registries (range: 0-2.5%). Ten non-melanoma skin cancers were self-reported but were included in the self-reported "no" numbers because cancer registries do not include these cancers. One cancer was reported as a metastatic cancer (not incident) and another as a "pre-cancer lesion"; these were both included in the self-report "no" category (n=7,965). The sensitivity of the registry search and linkage was 72% when using cancer self-reports as the gold-standard.

We examined the 82 self-reported cancers not found in the registries to determine whether these were in fact accurate cancer reports (Table 3). We successfully obtained medical records on 29 individuals; 13 of these confirmed the self-report (including correct date of diagnosis); 4 had no cancer in that time frame; 4 had non-melanoma skin cancer (they reported melanoma); 8 reported the correct cancer, but with an incorrect date of diagnosis (reported date was within the study period, but medical records date was outside the study period). There were also 17 self-reported cases that appeared to match cases in the registry, except that the dates did not match (reported date was within study period, but registry data show date outside the study period). Given that date of diagnosis is not always correctly self-reported, we assumed that the self-reported dates were incorrect (and registry dates correct), and did not attempt to obtain medical records for these 17 cases. We were unable to obtain further information on 36 self-reported cancers for different reasons: refusal to release medical records (n=8), no response from participant (n=13), no response from health care provider (n=1), no records available for the participant (n=1), illness (n=1), or no follow-up conducted (n=12).

The registry matching process detected 71 individuals with cancer (between baseline and 1998) for study participants who themselves did not report cancer on the follow-up questionnaire during that same period. Of the 71 individuals, 32 really did have a self-reported cancer, but because their reported diagnosis date was either missing or inaccurately self-reported, they had not been included in the matching, i.e., were excluded from the follow-up period (Table 4).

Table 3. Cancer Confirmation for Self-reported Cancers that Did Not Match to the Registries Between Baseline (1995–1996) and December 31, 1998

Status	Number	Percent of total	Percent with medical records
Confirmed self-reported cancer; no registry data available	13	15.9	44.8
Noncancer; responder or provider confirmed	4	4.9	13.8
Reported melanoma skin cancer; confirmed non-melanoma skin cancer	4	4.9	13.8
Reported cancer diagnosis date incorrect; medical records confirmed§	8	9.8	27.6
Reported cancer diagnosis date incorrect; registry confirmed (outside of appropriate time interval)	17	20.7	-
Unconfirmed self-reported cancer	36	43.9	-
Total	82	100	100

§Medical records indicate reported cancer was diagnosed before or after time period of interest.

Table 4. Registry Records Matched to NIH-AARP Cohort Pilot Members but No Cancers Were Reported by Those Individuals on the Follow-up Questionnaire Between Baseline (1995–1996) and December 31, 1998

Status	Number	Percent of total
Self-reported cancer did not include a diagnosis date	7	9.9
Self-reported cancer diagnosis date was pre-baseline	13	18.3
Self-reported cancer diagnosis date was post-1998	12	16.9
No self-reported cancer	39	54.9
Total	71	100

Table 5. Revised Table 2: Adjusting Self-reports with Medical Record Data and Inaccurate Diagnosis Dates (see text for more detail)

Self-reported (and confirmed)	Registry-reported		
	Yes	No	**Total**
Yes	239[§] (2.9%)	29[*] (0.4%)	268
No	39[§] (0.5%)	8019 (96.3%)	8058
Total	278	8048	8326

Sensitivity of registry = 239/268 = 89.2%
Specificity of registry = 8019/8058 = 99.5%

False negative rate = 29/268 = 10.8%

[*] Estimated cancers missed by registries: (44.8% x 36 = 16) + 13 (per table 3) = 29

[§] Based on table 4 adjustments

Therefore, these 32 individuals really should have been included as "yes" for self-reported cancers. With the adjusted values, only 39 registry cases were not self-reported on the questionnaire (Table 5).

Table 5 shows a revised Table 2 when accounting for results from medical record data and mismatches from missing data. We estimate the number of cases missed by the registries (false negatives) to be 29 (44.8% of 36 with no medical records, plus the 13 confirmed cases). From this table, and under the assumption that the confirmed self-report data represents true disease status, we estimate that 10.8% of cancers will be false negatives when using registry data only. Thus, registries included in this cohort have a sensitivity of 89.2% for all cancers.

Discussion

For the 12,000 pilot study participants of the NIH-AARP Diet and Health Study, observed incidence rates were similar to expected SEER rates between baseline and end of 1998. We attempted to retrieve medical records for those participants who self-reported a cancer diagnosis in the relevant time period but who did not have a cancer registry match. Based on these findings, we estimated that the registries included in the NIH-AARP cohort have close to 90% sensitivity with a 4-year lag period between the end of the period of interest and the reporting from cancer registries.

For a few cancers, observed case numbers were lower than expected. These differences were not statistically significant, suggesting that these variations are random and not likely to be due to reporting errors. However, for some site-specific cancers we had limited power to observe differences between observed and expected rates. For lung cancer, where the expected number of cases was 55 but only 41 cases were observed (and p-value 0.06) it is likely that this difference is due to the self-selection of "healthy" cohort participants, who tend to smoke less than the average American (and consequently have lower lung cancer). This is supported by data from a previous publication on this cohort where we showed that less than 11% of men were current smokers at baseline (1995-1996), which is lower than the national average smoking level at that time.[9] Overall, because it is unlikely that the cancer registries are biased at reporting some types of cancer better than others, our data suggest that cancer registries can be used to ascertain cases accurately for all cancer types.

Previous validation studies on case ascertainment have largely focused on the accuracy of self-reports, using cancer registries as the gold standard. Sensitivity rates of self-reports range between 61% and 93%,[1,3-4,15] and vary substantially by cancer site.[2] In this pilot study, we tested the sensitivity of the registry compared to self-reported cancers (with medical record confirmation). One of the advantages of using cancer registries for endpoint ascertainment is that cancer reporting varies little by site, unlike self-reporting.

Reporting lag is a main concern when using registry data. For this reason, follow-up time in this pilot study was only through the end of 1998. The registry search for the pilot study was conducted during 2003, and allowed for a minimum of a 4-year lag for cancers to be reported to the registries. These findings suggest that completeness of cancer reporting for the registries included in our cohort is close to the completeness of the SEER Program registries,[8] given a 4-year lag period.

We were unable to determine why 39 individuals who were matched to the cancer registries did not report a cancer on the follow-up questionnaire (Table 4). Previous validation studies of self-reports have shown that individuals often underreport cancer history.[3,15] In this study, the self-reporting false negative rate (39/278=14%; Table 5) was within the expected range, based on previous validation studies. Even so, mismatching of participants to the registries may increase numbers of false negatives. For around 5/6 of the cohort, we matched on Social Security numbers, which provides a high degree of accuracy for matching. Numbers were available for 27 of the 39 registry cancers not self-reported, and all had information on date of birth. Therefore, it is unlikely that all 39 cancers found in the registries are due to mismatching. Because it is likely that some of these 39 cancers are true cases, the calculated 89.2% sensitivity represents a slight underestimate of true sensitivity.

The percent of pilot study participants responding to the initial follow-up questionnaire, the postcard with an 800 number to complete the questionnaire, or follow-up phone interviews was 73. Because self-reported cancers were not very accurate among individuals who did not have a match to the cancer registries (44.8%; Table 3), inclusion of a disease follow-up questionnaire for the entire cohort would require medical record confirmation. Given response rates and medical record retrieval rates, adding a disease follow-up questionnaire to our cohort would decrease false negatives by no more than 3% (56% follow-up questionnaire response (without phone interview) x 54% medical record retrieval x 10% false negatives). Therefore, adding a disease follow-up questionnaire and medical record confirmation to this cohort would increase sensitivity to no more than 93%.

Incomplete case ascertainment when specificity is perfect (or close to 100%) does not bias the risk ratio when the misclassification is nondifferential,[16] but it will result in loss of power. In this cohort, the expected number of cancer cases over a short follow-up period is very high, given the size of the study.[9] Therefore, the statistical power to detect important associations in this cohort should be minimally compromised by a sensitivity that is approximately 90%. However, site-specific sensitivities may differ somewhat from that figure.

The NIH-AARP cohort was designed to include only individuals living in cancer registry states. Over time, participants may move out of the cohort region, which would result in undetected cancer cases. However, given that during 9 years of follow-up only 2.5% (288/11,404) of surviving pilot study subjects moved out of cohort regions, it is unlikely that migration would result in a substantial number of missing cancer cases.

Conclusion

We estimate that by matching NIH-AARP cohort participants to cancer registries, about 90% of all cancer cases will be detected in this cohort when allowing for a 4-year lag between cancer diagnosis date and matching year. The accuracy of cancer endpoint ascertainment through cancer registries in this cohort is comparable to that of other cohort studies that rely on self-reporting (with medical record confirmation) for case ascertainment, and should provide accurate cancer numbers for analyses in this cohort.

References

1. Bergmann MM, Calle EE, Mervis CA, Miracle-McMahill HL, Thun MJ, Heath CW. Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries. *Am J Epidemiol.* 1998;147(6):556-62.
2. Parikh-Patel A, Allen M, Wright WE. Validation of self-reported cancers in the California Teachers Study. *Am J Epidemiol.* 2003;157(6):539-45.
3. Desai MM, Bruce ML, Desai RA, Druss BG. Validity of self-reported cancer history: A comparison of health interview data and cancer registry records. *Am J Epidemiol.* 2001;153(3):299-306.
4. Schrijvers CT, Stronks K, van de Mheen DH, Coebergh JW, Mackenbach JP. Validation of cancer prevalence data from a postal survey by comparison with cancer registry records. *Am J Epidemiol.* 1994;139(4):408-14.
5. Berthier F, Grosclaude P, Bocquet H, Faliu B, Cayla F, Machelard-Roumagnac M. Prevalence of cancer in the elderly: discrepancies between self-reported and registry data. *Br J Cancer.* 1997;75(3):445-7.
6. NAACCR. *Standards for Completeness, Quality, Analysis, and Management of Data,* Vol. 3. North American Association of Central Cancer Registries; 2002.
7. SEER. *Data Quality.* Bethesda, MD: Vol. 2003; 2003.
8. Clegg LX, Feuer EJ, Midthune DN, Fay MP, Hankey BF. Impact of reporting delay and reporting error on cancer incidence rates and trends. *J Natl Cancer Inst.* 2002;94(20):1537-45.
9. Schatzkin A, Subar AF, Thompson FE, Harlan LC, Tangrea J, Hollenbeck AR, et al. Design and serendipity in establishing a large cohort with wide dietary intake distributions : The National Institutes of Health-American Association of Retired Persons Diet and Health Study. *Am J Epidemiol.* 2001;154(12):1119-25.
10. Fellegi IP, Sunter AB. A theory for record linkage. *J Am Stat Assoc.* 1969;64:1183-1210.
11. Hermansen SW. *Fuzzy Key Linkage: Robust Data Mining Methods for Real Databases.* Firenze, Italy: SAS European Users Group International (SeUGI); 2000.
12. Monson RR. Analysis of relative survival and proportional mortality. *Comput Biomed Res.* 1974;7(4):325-32.
13. SEER*Stat software version 4.2. Bethesda, MD: National Cancer Institute, DCCPS, Surveillance Research Program; 2002.
14. Surveillance, Epidemiology, and End Results (SEER) Program SEER*Stat Database: Incidence - SEER 9 Regs Public-Use (1973-1999). Bethesda, MD: National Cancer Institute, DCCPS, Surveillance Research Program; 2001.
15. Paganini-Hill A, Chao A. Accuracy of recall of hip fracture, heart attack, and cancer: A comparison of postal survey data and medical records. *Am J Epidemiol.* 1993;138(2):101-6.
16. Rothman KJ, Greenland S. *Modern Epidemiology,* 2nd ed. Philadelphia, PA: Lippincott-Raven Publishers; 1998.

Acknowledgements

Cancer incidence data from the Atlanta metropolitan area were collected by the Georgia Center for Cancer Statistics, Department of Epidemiology, Rollins School of Public Health, Emory University. Cancer incidence data from California were collected by the California Department of Health Services, Cancer Surveillance Section. Cancer incidence data from the Detroit metropolitan area were collected by the Michigan Cancer Surveillance Program, Community Health Administration, State of Michigan. The Florida cancer incidence data used in this report were collected by the Florida Cancer Data System under contract to the Department of Health (DOH). The views expressed herein are solely those of the authors and do not necessarily reflect those of the contractor or DOH. Cancer incidence data from Louisiana were collected by the Louisiana Tumor Registry, Louisiana State University Medical Center in New Orleans. Cancer incidence data from New Jersey were collected by the New Jersey State Cancer Registry, Cancer Epidemiology Services, New Jersey State Department of Health and Senior Services. Cancer incidence data from North Carolina were collected by the North Carolina Central Cancer Registry. Cancer incidence data from Pennsylvania were supplied by the Division of Health Statistics and Research, Pennsylvania Department of Health, Harrisburg, Pennsylvania. The Pennsylvania Department of Health specifically disclaims responsibility for any analyses, interpretations, or conclusions.